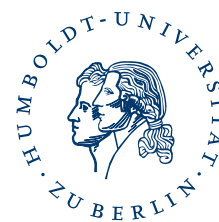


HUMBOLDT-UNIVERSITÄT ZU BERLIN



# Domain-Specific CLIR of English, German and Russian Using Fusion and Subject Metadata for Query Expansion

Vivien Petras<sup>1</sup>, Fredric Gey<sup>2</sup>, and Ray R. Larson<sup>1</sup>

<sup>1</sup>*School of Information Management and Systems  
University of California, Berkeley, CA 94720 USA  
[vivienp@sim.s.berkeley.edu](mailto:vivienp@sim.s.berkeley.edu) / [ray@sim.s.berkeley.edu](mailto:ray@sim.s.berkeley.edu)*

<sup>2</sup>*UC Data Archive & Technical Assistance (UC DATA)  
University of California, Berkeley, CA 94720 USA  
[gey@berkeley.edu](mailto:gey@berkeley.edu)*

This is an author's accepted manuscript version of a conference paper published in *International Conference of the Cross-Language Evaluation Forum for European Languages, CLEF 2005: Accessing Multilingual Information Repositories* within the Springer Lecture Notes in Computer Science book series (LNCS, volume 4022).

The final publisher's version is available online at:  
[https://doi.org/10.1007/11878773\\_27](https://doi.org/10.1007/11878773_27)

# Domain-Specific CLIR of English, German and Russian Using Fusion and Subject Metadata for Query Expansion

Vivien Petras<sup>1</sup>, Fredric Gey<sup>2</sup> and Ray R. Larson<sup>1</sup>

<sup>1</sup>School of Information Management and Systems  
University of California, Berkeley, CA 94720 USA  
{vivienp, ray}@sims.berkeley.edu  
<sup>2</sup>UC Data Archive & Technical Assistance (UC DATA)  
University of California, Berkeley, CA 94720 USA  
gey@berkeley.edu

**Abstract.** This paper describes the combined submissions of the Berkeley group for the domain-specific track at CLEF 2005. The data fusion technique being tested is the fusion of multiple probabilistic searches against different XML components using both Logistic Regression (LR) algorithms and a version of the Okapi BM-25 algorithm. We also combine multiple translations of queries in cross-language searching. The second technique analyzed is query enhancement with domain-specific metadata (thesaurus terms). We describe our technique of Entry Vocabulary Modules, which associates query words with thesaurus terms and suggest its use for monolingual as well as bilingual retrieval. Different weighting and merging schemes for adding keywords to queries as well as translation techniques are described.

## 1 Introduction

For CLEF 2005, the Berkeley group split into two groups (Berkeley 1 and Berkeley 2). Berkeley 1 focused on data fusion techniques whereas Berkeley 2 focused on query expansion techniques using subject metadata. The groups used different probabilistic algorithms and retrieval systems. In this paper, we will report all our results for the domain-specific track but concentrate on describing Berkeley 2's retrieval techniques whereas our paper for the GeoCLEF track will mainly describe Berkeley 1's retrieval techniques (see [9]).

### 1.1 Fusion

Fusion is a retrieval technique based on the assumption that several different information retrieval systems will retrieve more relevant results than a single retrieval algorithm alone. Lee [10] found that result sets from different retrieval algorithms show similar relevant documents and different non-relevant documents providing criteria for finding relevant documents in a merged set by emphasizing documents found more than once and downweighting documents that are unique to each algorithm.

In [8], the Berkeley 1 group experimented with the fusion of a logistic regression algorithm and the OKAPI BM-25 algorithm. A combination of these two algorithms is also used in the CLEF 2005 experiments.

The search results were combined using the CombMNZ data fusion algorithm developed by Shaw and Fox [16]. The CombMNZ algorithm merges result lists, normal-

izing the scores in each list and increasing scores for items based on the number of result lists that they appear in, while penalizing items that appear in only a single list.

## **1.2 Query Expansion Using Subject Metadata**

Query expansion has been researched in the information retrieval field for a long time [4]. However, automatic query expansion has been mostly discussed in the context of blind feedback or highly evolved expert systems [e.g. 5]. Thesauri or subject metadata in general are mainly used for manual or interactive query expansion (for an overview, see [17]), but authors report mixed results [6,18] when comparing those techniques to free-text search.

For CLEF 2005, Berkeley's group 2 experimented with Entry Vocabulary Modules (EVMs) to automatically enhance queries with subject metadata terms or to replace query terms with them.

The technique of Entry Vocabulary Modules was designed to serve as an interface between the query vocabulary of the searcher (natural language) and the controlled vocabulary entries of a database. Given any search word or phrase, it will suggest controlled vocabulary terms that represent the concept of the search. A searcher can use these terms to append to his or her query or to substitute his or her own query terms with those controlled vocabulary terms in the hope of achieving a more precise and complete retrieval.

This technique can be used for automatic query expansion if the selection of EVM-suggested thesaurus terms for appending to the query is predetermined (e.g. a number of top-ranked EVM-suggested terms are automatically added to the query).

## **2 CLEF Domain-Specific Collections**

The GIRT collection (German Indexing and Retrieval Test database) consists of 151,319 documents (parallel in English and German) containing titles, abstracts and thesaurus terms in the social science domain. The GIRT thesaurus terms are assigned from the Thesaurus for the Social Sciences [15] and are provided in German, English and Russian. For a detailed description of GIRT and its uses, see [7].

The English GIRT collection contains only 26,058 abstracts (ca. one out of six records) whereas the German collection contains 145,941 - providing an abstract for almost all documents. Consequently, the German collection contains more terms per record to search on. The English corpus has 1,535,445 controlled vocabulary entries (7064 unique phrases) and the German corpus has 1,535,582 controlled vocabulary entries (7154 unique phrases) assigned. On average, 10 controlled vocabulary terms / phrases are appended to each document.

Controlled vocabulary terms are not uniformly distributed. Most thesaurus terms occur less than a 100 times, but 307 occur more than 1,000 times and the most frequent one, "Bundesrepublik Deutschland", occurs 60,955 times.

The Russian Social Science Corpus consists of 94,581 documents containing titles (for all documents) abstracts (for 47,130 documents or 50% of the collection). Unfortunately for this collection, only 12% of the collection (11,403 documents) have controlled-vocabulary thesaurus terms assigned.

### 3 Entry Vocabulary Modules

An Entry Vocabulary Module is a dictionary of associations between terms in titles and abstracts in documents and the controlled vocabulary terms associated with the document. If title/abstract words and thesaurus terms co-occur with a higher than random frequency, there exists a likelihood that they are associated. A likelihood ratio statistic is used to measure the association between any natural language term and a controlled vocabulary term. Each pair is assigned an association weight (rank) representing the strength of their association. The higher the rank, the more a thesaurus term represents the concept represented by the document word. The methodology of constructing Entry Vocabulary Modules has been described in detail in [13].

Once an Entry Vocabulary Module is constructed and a table of associations and their weights exist, we can look up a word in the dictionary and find its most highly associated thesaurus term. This is how we find thesaurus terms to associate with the GIRT queries. After experimenting with looking up query title and description words, we found that query title words are sufficient to find relevant thesaurus terms. For all CLEF 2005 experiments, only query title words (after stopword removal) were used for thesaurus term look-up.

If more than one word appears in the query title, we need to merge the results from the thesaurus term look-ups to receive a list of terms for the query as a whole. We experimented with two merging strategies.

For absolute rank merging, an absolute rank for each thesaurus term is calculated by adding the association weights if it is associated with several title words. The five thesaurus terms with the highest rank are then added to the query. The pitfall of this merging strategy is that some association pairs have such high weights that other important query word – thesaurus term combinations will be ranked lower no matter what. To avoid this problem, we also tested a round robin merging strategy: for each query word, we looked up the two highest ranked thesaurus terms and added them to the query.

Table 1 shows 2 examples for the different merging strategies and their advantages and disadvantages. For query 138, the first two thesaurus terms in the round robin strategy are highly associated with “insolvent”, the second two with “companies”. As one can see in the absolute rank strategy, the thesaurus terms for “companies” seem to ‘overpower’ the ones for “insolvent”. Sometimes however, this strategy is prone to errors as topic 143 proves. The words looked up in the EVM are “smoking” and “giving”, which is misleading. The absolute rank strategy performs better in this case.

**Table 1.** Comparison of absolute rank and round robin merging for 2 queries

Query 138: Insolvent Companies		Query 143: Giving up Smoking	
<i>Absolute rank merging</i>	<i>Round robin merging</i>	<i>Absolute rank merging</i>	<i>Round robin merging</i>
enterprise	liquidity	smoking	donation
firm	indebtedness	tobacco consumption	social relations
medium-sized firm	enterprise	tobacco	smoking
small-scale business	firm	behavior modification	tobacco consumption
flotation		behavior therapy	

For German with its compounds (“Unternehmensinsolvenzen” instead of “Insolvent Companies” for topic 138), the round robin strategy sometimes only adds two instead of five thesaurus terms to the query, the ranking otherwise being equal to the absolute rank strategy.

For a more in-depth explanation of EVMs and the merging strategies, see our CLEF2005 working paper [12].

## **4 Retrieval Techniques**

### **4.1 Berkeley 1 – Fusion**

For both the monolingual and bilingual tasks we indexed the documents using the Cheshire II system. The logistic regression algorithm used was the Berkeley TREC-3 algorithm [3], the OKAPI BM-25 algorithm is based on Robertson [14]. The document index entries and queries were stemmed using the Snowball stemmer. Text indexes were created for separate XML elements (such as document titles or dates) as well as for the entire document. The techniques and algorithms used for the domain-specific task were essentially identical to those that we used for the GeoCLEF task and are described in the paper for that track (see [9] for more detail).

### **4.2 Berkeley 2 – EVM Query Expansion**

In all its CLEF submissions, the Berkeley 2 group used a document ranking algorithm based on logistic regression first used in the TREC-2 conference [1]. For all runs, we used stopword lists to remove very common words from collections and queries as well as an implementation of the Muscat stemmer for both English and German and the Snowball stemmer for Russian. For German runs, we used a decompounding procedure developed and described by Aitao Chen [2], which has been shown to improve retrieval results. The decompounding procedure looks up document and query words in a base dictionary and splits compounds when found. As a general procedure, we also use Aitao Chen’s blind feedback algorithm [2] in every run. It selects the top 30 ranked terms from the top 20 ranked documents from the initial search to merge with the original query.

Thus, the sequence for processing for retrieval is: query → stopword removal → (decompounding) → stemming → ranking → blind feedback.

## **5 Retrieval Results – Fusion**

The data fusion experiment results did not have a very good performance. Relative to our German and English results, the Russian results look fairly good (we suspect that this may be due to the smaller number of participants). Among the beneficial techniques used in the better-performing Berkeley 2 group runs are 1) query expansion from the thesaurus, 2) automatic decompounding of German words and 3) application of blind relevance feedback. The official submitted runs can be considered preliminary baselines that, we hope, will be improved upon in the future.

The primary approach used by the Berkeley 1 group for query processing is quite similar to that described above, however, no decompounding or blind feedback steps

were used, and the ranking algorithms were different, and included multiple ranked sets of results that were then merged using data fusion methods for the final submitted results.

Table 2 shows the average precision for the Berkeley 1 group's submitted runs for the Monolingual tasks. In the monolingual runs, the topic description and title were combined and searched using the TREC3 logistic regression algorithm, and the Okapi BM-25 algorithm. The results of these two searches were then combined using the CombMNZ algorithm. As can be seen by comparison with the results reported by the Berkeley 2 group, the results were not impressive for this task.

**Table 2.** Average precision scores for Berkeley 1 monolingual title + description runs for German, English and Russian

Run	BERK1MLDE	BERK1MLEN	BERK1MLRU
Avg. precision	0.2314	0.3291	0.2409

**Table 3.** Average precision scores for Berkeley 1 bilingual title + description runs

run	BERK1 BLDEEN	BERK1 BLDERU	BERK1 BLENDE	BERK1 BLENRU	BERK1 BLRUDE	BERK1 BLRUEN
Languages	German -> English	German-> Russian	English-> German	English-> Russian	Russian-> German->	Russian-> English
Translators	BabelFish L&H	Prompt	BabelFish L&H	Prompt	Prompt	BabelFish Prompt
Avg. Precision	0.2398	0.1717	0.1477	0.1364	0.1687	0.2358

Table 3. shows the average precision of the bilingual runs for the Berkeley 1 group. Once again, comparison with Berkeley 2 results for the corresponding tasks shows a significant gap in the performance of the fusion methods when compared to their methods (including decompounding of German Terms, the TREC2 logistic regression algorithm and blind feedback).

**Table 4.** Average precision scores for Berkeley 1 multilingual tasks

Run	BERK1MUDEALL	BERK1MUENALL	BERK1MURUALL
Languages	German-> German, English, Russian	English-> German, English, Russian	Russian-> German, English, Russian
Translators	BabelFish L&H Prompt	BabelFish L&H Prompt	BabelFish Prompt
Avg. Precision	0.0294	0.0346	0.0532

Table 4. shows the results for the Berkeley 1 multilingual runs (again using title and description). The results are very low (especially when compared to the Berkeley 2 group Monolingual and Bilingual runs). However these were the top-ranked runs for

the DS Multilingual task (of course, they are also, apparently, the *only* submissions for the DS Multilingual task).

It is worth noting that the Berkeley 1 group ran some post-CLEF tests (to verify that the results obtained were not the result of system errors, but instead were the result of the behavior of the fusion operation and the retrieval algorithms used in the CLIR tasks. The tests involved using the TREC2 logistic regression algorithm with blind feedback in place of the TREC3 algorithm while using the same parsing and stemming techniques used the runs reported above, but not using data fusion methods or OKAPI for ranking. The results of these (monolingual only runs) showed considerable improvement for all languages for monolingual retrieval compared to the fusion approach, and were very close to the Berkeley 2 results for title+description English and Russian (0.4472 and 0.2979 Average Precision, respectively). For monolingual German, our post-result was 0.2769 Average Precision. These results highlight the very important effects of using query expansion, and decompounding of German words on performance (as well as choosing the best single algorithm for the task). We believe, however that there may have been some anomalies in the application of the CombMNZ fusion algorithm for some of our tasks, so we intend to do some further investigation of the results in planning for next year's tasks.

## 6 Retrieval Results – Query Expansion

For more experiments and an in-depth analysis, see our CLEF2005 working paper [16].

### 6.1 Monolingual Retrieval

For monolingual retrieval, we experimented with three query expansion strategies:

- adding five thesaurus terms retrieved with the EVM absolute rank merging from query title words;
- adding five thesaurus terms from the absolute rank merging strategy (using only query title words) but removing all thesaurus terms from the dictionary that occurred more than a 1,000 times in the document collection, thereby hoping to remove thesaurus terms that would not discriminate effectively;
- adding two thesaurus terms retrieved from the EVM for each query title word using the round robin merging strategy.

For every expansion strategy, we analyze one run where the thesaurus terms are downweighted and one where they are treated as equally important part of the query.

**6.1.1 German.** As the following table 5 shows, query expansion always improves over the baseline run of title+description if the expanded part is downweighted. If the thesaurus terms are not downweighted, only the round robin strategy improves over the baseline run. However, this case is also the dominating strategy, not only improving the baseline by 13% but also improving on the downweighted strategy and on the other merging strategies.

**Table 5.** Average precision scores for title + description German Monolingual runs

Run	TD baseline	ABS HW	ABS	ABS -1000 HW	ABS -1000	RR HW	RR
Official run	BK2G MLGG1	BK2G MLGG2		BK2G MLGG3		BK2G MLGG4	
Avg. precision	0.4547	0.4733	0.4369	0.4595	0.3866	0.4936	<i>0.5144</i>

ABS absolute rank strategy

ABS -1000 absolute rank strategy omitting thesaurus terms that occur more than 1000times in the collection

RR round robin merging

HW expanded thesaurus terms are downweighted by half in this run

Comparing precision on a query-by-query basis, it becomes clear that downweighting clearly dominates for the absolute rank strategies, whereas not downweighting equally dominates for the round robin strategy although the average precision scores are much closer. In 18 of 25 queries, absolute rank merging with downweighting had a better precision than the not downweighted absolute rank strategy, for the absolute rank – 1000 strategy, downweighting achieved a better result in 20 cases. For round robin, not downweighting turned out to be better in 17 of 25 cases compared to downweighting.

**6.1.2 English.** As table 6 shows, query expansion with EVM suggested thesaurus terms is not as successful for English monolingual retrieval. However, the trend remains the same as in German monolingual retrieval. The round robin strategy without downweighting is still the dominating strategy, improving on the baseline by 6%. For the absolute rank strategies, downweighting works better, although they don't improve on the baseline.

**Table 6.** Average precision scores for title + description English Monolingual runs

Run	TD baseline	ABS HW	ABS	ABS -1000 HW	ABS -1000	RR HW	RR
Official run	BK2G MLEE1	BK2G MLEE2		BK2G MLEE3			
Avg. precision	0.4531	0.4149	0.3462	0.4125	0.3092	0.4697	<i>0.4818</i>

The difference between downweighting or not is more pronounced when looking at the results on a query-by-query basis: in 21 out of 25 cases downweighting is better for the absolute rank strategy and in 20 of 25 cases for the absolute rank –1000 strategy. Not downweighting works better for round robin merging in 14 out of the 25 cases.



In some cases, the absolute strategy seems to make things much worse. This is because it adds thesaurus terms that are too general. But even the round robin strategy doesn't seem to improve precision as much as in German monolingual retrieval. Ironically, it seems that the unique characteristics of the German language (compounds) help in suggesting thesaurus terms that are not only more on the mark but are also compounds themselves retrieving more relevant documents. For example, the thesaurus term *way of life* translates to *Lebensweise* in German. Whereas for English, the retrieval system will look for documents containing "way" and "life" (very general!), the retrieval system will look for "Lebensweise" in German, which is much more precise.

However, it also cannot be overlooked that the English collection contains less text (fewer abstracts) than the German collection to search on. It might be that the added thesaurus terms skew search results in that they take away weight from the free-text search terms ranking documents containing the thesaurus terms higher than ones containing the free-text search terms. This would explain the greater improvement of the downweighting strategies for absolute rank merging as compared to German (precision increases by 20% and 33% for ABS and ABS -1000 in English, whereas only by 8% and 19% in German) and the smaller improvement of not downweighting for round robin (2.5% in English vs. 4% in German).

## 6.2 Bilingual Retrieval

For bilingual retrieval, we experimented with query expansion and query reformulation using EVMs in addition to query translation. Three translation techniques are compared:

1. Machine translation. We used a combination of the Systran translator (<http://babelfish.altavista.com/>) and the L & H Power Translator.
2. Thesaurus matching. Words and phrases from the query are looked up in the thesaurus with a fuzzy-matching algorithm and if a matching thesaurus term in the query language is found, the equivalent thesaurus term in the target language is used. See [11] for a more detailed description.
3. EVM. The query title words were submitted to the query language EVM and the round robin merging technique was used to retrieve thesaurus terms. The thesaurus terms in the query language were then replaced by the thesaurus terms in the target language. The query was then reformulated using only thesaurus terms.

We have combined translation techniques by submitting the translated output from the different methods in one and the same run. This increases the number of query words and the danger of introducing more non-discriminating search terms as well as favoring easy to translate terms (they most likely to occur in all methods), but for CLEF, this strategy has worked successfully in previous years. Combining translation methods helps with hard to translate words (higher chance of one method getting it right) and reduces the risk of mis-translation. Table 7 compares combination runs for German-English and English-German retrieval.

For German-English, a combination of all three techniques is clearly the dominating strategy – it seems that adding more words describing the same concept generally improves the precision instead of adding too many non-discriminating terms. It is also worth mentioning that all combination runs perform better than machine translation alone (avg. precision 0.3917), even if one combines thesaurus matching and EVM terms only. In fact, even though lower in precision, this combination performs better in 13 out of 25 cases compared to both the machine translation – thesaurus matching and the machine translation – EVM pairs; a worthy competitor to the commercial translation solutions.

**Table 7.** Bilingual retrieval combining translation methods

	Machine Translation + Thesaurus Matching	Machine Translation + EVM thesaurus terms	Thesaurus Matching + EVM thesaurus terms	Machine Translation + Thesaurus Matching + EVM thesaurus terms
<i>German-English</i>				
Avg. precision	0.4514	0.4566	0.4346	0.4803
<i>English-German</i>				
Avg. precision	0.4201	0.4059	0.4254	0.4374

For English-German retrieval, all combination runs seem to perform similarly. However, once again, they clearly outperform machine translation alone (avg. precision 0.3532). Of course, not all combinations work equally well for each query and, sometimes, one translation technique alone works much better.

### 6.3 Summary

Expanding a query with terms from a thesaurus is like asking an information expert to translate your search strategy into the search language of the database, hopefully providing better search terms than the original search statement. The information expert for this set of experiments is an association dictionary of thesaurus terms and free-text words from titles and abstracts from the collection. Based on title words from the query, thesaurus terms that are highly associated with those words are suggested. Two merging strategies have been tested: absolute rank merging, based on all title words as a set and round robin merging, which suggests two thesaurus terms for each individual query word.

For monolingual retrieval, query expansion with EVM suggested thesaurus terms improves over the baseline of title + description submission by 13% (German) and 6% (English), respectively. Downweighting the added terms performs better for absolute rank but not for the round robin merging. For German, submitting only thesaurus terms (replacing the original query) decreases the average precision over 25 cases, but achieves better precision in 12 individual cases.

For bilingual retrieval, using the thesaurus for translation works surprisingly well. Just using thesaurus terms for the query submission works almost as well as machine translation. Although average precision decreases (9% for English-German and 15% for German-English), EVM suggested thesaurus terms perform better in one third of the queries. A combination of two thesaurus techniques (EVM and thesaurus matching) outperforms machine translation. The combination of machine translation, thesaurus matching and EVM suggested terms outperforms all other strategies.

It has been shown that EVM suggested terms can provide the impact to raise precision for a query – if they are high quality search terms. High quality search terms are those that provide discriminating search power (they occur mostly in relevant documents), describe the information need exactly and, ideally, add new terms to the query. Added terms that are too vague will almost always degrade the performance.

## 7 Retrieval Results - Russian

The Berkeley 2 group results are summarized by topic in the following table with comparison to overall precision. The highlighted columns are the median performances for monolingual and cross-language IR while the final row is precision averaged over all 25 topics:

**Table 8.** Berkeley 2 Russian monolingual and bilingual results

Topic	Best Mono	Med Mono	BK2M LRU1	BK2M LRU2	Best CLIR	Med CLIR	BK2B LER1	BK2B LGR1
126	0.5437	0.2004	0.5437	0.2083	0.5182	0.4119	0.421	0.5182
127	0.9036	0.8295	0.9036	0.8789	0.8691	0.6872	0.8691	0.7559
128	0.7085	0.2613	0.2783	0.1973	0.3793	0.2374	0.2594	0.3793
129	0.0596	0.0279	0.0596	0.0095	0.0021	0	0.0021	0.0011
130	0.1227	0.0143	0.0801	0.026	0.0597	0.0061	0.0025	0.0061
131	1	0.0005	1	0.5089	0.5294	0.0976	0.5294	0.2976
132	0.125	0.027	0.125	0.0312	0.304	0.125	0.125	0.1
133	0.1791	0.0606	0.1716	0.1152	0.4643	0.1071	0.3915	0.4643
134	0.3917	0.0992	0.1024	0.0959	0.0913	0.02	0.0913	0.0607
135	0.534	0.1463	0.1419	0.534	0.1876	0.0801	0.1876	0.0257
136	0.6905	0.5087	0.585	0.4324	0.1109	0.022	0.1109	0.1002
137	0.287	0.1797	0.287	0.1855	0.191	0.1114	0.1555	0.191
138	0.5313	0.4702	0.4727	0.3337	0.177	0.0432	0.0432	0.177
139	0.616	0.4282	0.3966	0.4223	0.5145	0.2241	0.2294	0.5145
140	0.0503	0.0368	0.0292	0.0342	0.0358	0.0271	0.0255	0.0271
141	0.2847	0.0454	0.0539	0.2847	0.2086	0.1933	0.1933	0.1344
142	0.7698	0.3085	0.3731	0.2439	0.2886	0.0678	0.0136	0.2886
143	1	0.2667	1	0.45	1	0.7381	0.0094	1
144	0.0402	0.0089	0.0056	0.0091	0.027	0.0137	0.0065	0.0137
145	0.6553	0.5809	0.5335	0.2058	0.6821	0.5949	0.5949	0.6821
146	0.0435	0.0197	0.004	0.0091	0	0	0	0
147	0.125	0	0	0.125	0.0016	0	0.0011	0

148	0.3939	0.2492	0.2405	0.3587	0.1618	0.0639	0.1618	0.0551
149	0.2066	0.0111	0.2066	0.1734	0.088	0.0257	0.088	0.0257
150	0	0	0	0	0.0178	0.0139	0.0139	0.0102
<b>Avg</b>	<b>0.3887</b>	<b>0.1832</b>	<b>0.3038</b>	<b>0.2349</b>	<b>0.2557</b>	<b>0.14</b>	<b>0.181</b>	<b>0.2331</b>

The first monolingual Russian run (**BK2MLRU1**) and the two bilingual runs (**BK2BLER1**, **BK2BLER2**) were made using the required Title and Description (TD) fields. The second monolingual run (**BK2MLRU2**) used the Title, Description and Narrative (TDN) fields. The TD run (**BK2MLRU1**) achieved overall mean average precision of 0.304 with 9 best-of-topic results out of the 25 topics. Interestingly, the TD run performed 30 percent higher than the TDN monolingual run (**BK2MLRU2**) which had an average precision of only 0.235. We speculate that this is because over half the documents in the collection only have a <TITLE> field and not a <TEXT> field. Topic 150 Поведение во время телепередач (Television Behaviour) retrieved zero relevant documents from all DS monolingual runs, while bilingual runs to the Russian found only two relevant document with best average precision of 0.0178.

The German-Russian bilingual run **BK2BLGR1** (MAP of 0.233) performed 29% better than the English-German run **BK2BLER1** (MAP of 0.181). Much of this difference can be attributed to topic 143 Отказ от курения (Giving up Smoking) where the German translation seems to have been more accurate than the English one. The German-->Russian precision for topic 143 was 1.0 while the English-->Russian precision was 0.0094.

We believe we achieved our goal of providing a baseline performance for the Russian domain-specific collection of CLEF. We believe our results provide a foundation from which more sophisticated experiments can be developed which leverage the controlled vocabulary indexing of the CLEF DS collections. For the future of CLEF domain-specific Russian to be interesting and successful, substantially more documents will need to have indexing keywords assigned to the documents – 12 % is simply not enough to perform meaningful experiments on the utility of controlled vocabulary.

## 8 Acknowledgements

Thanks to Aitao Chen for implementing and permitting the use of the logistic regression formula for probabilistic information retrieval as well as German decompounding and blind feedback in his MULIR retrieval system.

## References

- 1 Chen, A., Cooper, W. and F. Gey (1994). Full text retrieval based on probabilistic equations with coefficients fitted by logistic regression. In: D.K. Harman (Ed.), The Second Text Retrieval Conference (TREC-2): 57-66, March 1994.
- 2 Chen, A. and F. Gey (2004). "Multilingual Information Retrieval Using Machine Translation, Relevance Feedback and Decompounding." Information Retrieval 7(1-2): 149-182.

- 3 Cooper, W. S., A. Chen, et al. (1994). Experiments in the probabilistic retrieval of full text documents. Third Text Retrieval Conference (TREC-3), Gaithersburg, MD, National Institute of Standards and Technology Special Publication 500-225
- 4 Efthimiadis, E. N. (1996). Query Expansion. In Annual Review of Information Systems and Technology (ARIST), edited by M. E. Williams. Medford, NJ: Information Today.
- 5 Gauch, S. and J. B. Smith (1993). "An expert system for automatic query reformation." *Journal of the American Society for Information Science* 44(3): 124-36.
- 6 Jones, S. (1995). "Interactive thesaurus navigation: intelligence rules OK?" *Journal of the American Society for Information Science* 46(1): 52-9.
- 7 Kluck, M. (2003). The GIRT Data in the Evaluation of CLIR Systems - from 1997 Until 2003. 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003. C. A. Peters. Trondheim, Norway, August 21-22, 2003, LNCS 3237, Springer 2004: 376-390.
- 8 Larson, R. R. (2005). A fusion approach to XML structured document retrieval. *Information Retrieval* 8: 601-629.
- 9 Larson, R.R., Gey, F. & Petras, V. (2005). Berkeley at GeoCLEF: Logistic Regression and Fusion for Geographic Information Retrieval. This Volume.
- 10 Lee, J. H. (1997). Analyses of multiple evidence combination. Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 27-31, 1997, Philadelphia, ACM: 267--276.
- 11 Petras, V., N. Perelman, et al. (2003). UC Berkeley at CLEF 2003 -- Russian Language Experiments and Domain-Specific Cross-Language Retrieval. 4th Workshop of the Cross-Language Evaluation Forum, CLEF 2003. Trondheim, Norway, August 21-22, 2003, LNCS 3237, Springer 2004: 401-411.
- 12 Petras, V. (2005). How One Word Can Make all the Difference - Using Subject Metadata for Automatic Query Expansion and Reformulation. Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria [http://www.clef-campaign.org/2005/working\\_notes/workingnotes2005/petras05.pdf](http://www.clef-campaign.org/2005/working_notes/workingnotes2005/petras05.pdf)
- 13 Plaunt, C., and B. A. Norgard (1998). An Association-Based Method for Automatic Indexing with Controlled Vocabulary. *Journal of the American Society for Information Science* 49, no. 10 (1998), pp. 888-902.
- 14 Robertson, S. E. and S. Walker (1997). On relevance weights with little relevance information. Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval, ACM Press: 16-24.
- 15 Schott, H. (2000). Thesaurus for the Social Sciences. [Vol. 1:] German-English. [Vol. 2:] English-German. IZ Sozialwissenschaften Bonn, 2000.
- 16 Shaw, J. A. and E. A. Fox (1994). Combination of multiple searches. Proceedings of the 2nd Text REtrieval Conference (TREC-2), National Institute of Standards and Technology Special Publication 500-215: 243--252.
- 17 Shiri, A. A., C. Revie, et al. (2002). "Thesaurus-enhanced search interfaces." *Journal of Information Science* 28(2): 111-22.
- 18 Sihvonen, A. and P. Vakkari (2004). "Subject knowledge improves interactive query expansion assisted by a thesaurus." *Journal of Documentation* 60(6): 673-690.